

reprint

from

Department of Educational and Psychological
Research

School of Education, Malmö, Sweden

No. 119

AN ENGLISH DIGEST: SELF-CONFRONTATION
VIA CLOSED-CIRCUIT TELEVISION IN TEACHER
TRAINING TOGETHER WITH RECOMMENDA-
TIONS FOR FURTHER RESEARCH

Bernhard Bierschenk

AN ENGLISH DIGEST:
SELF-CONFRONTATION VIA CLOSED-
CIRCUIT TELEVISION IN TEACHER
TRAINING TOGETHER WITH RECOM-
MENDATIONS FOR FURTHER
RESEARCH

31. Problems

One of the goals of Swedish teacher training in pedagogics according to the directives of the Swedish National Board of Education (SÖ, Klasslärarutbildningen, studieplaner, 1968, s. 393) is

“to combine with other aspects of the training in furthering the personal development and self-knowledge of the aspirant teachers and providing them with a vocational training which, given their individual qualities, will equip them for undertaking the responsibility of educating and teaching the grades for which they are being trained”.

It has long been the policy within teacher training to let the student teachers discover for themselves, during their period of supervised teaching practice, the way in which their own behavior influences the teaching process. A well-known phenomenon within the psychology of perception is that the individual's structure of perception and evaluation leads to interpretations of objects and situations that are specific for that individual. One of the aims of the traditional tutorial system is to help the student teachers find out about themselves and about the relationships that exist between the student teacher and his pupils. But the difficulties experienced by both the tutor and the student teacher in recalling “*exactly*” what happened in the practice teaching situation obstruct the fulfillment of this aim.

In recent years, externally mediated self-confrontation by means of closed-circuit television and video-recording (CCTV/VR) has become a popular technique of confrontation. A large number of reports and articles, extremely varied in quality, have been published in many different journals. In addition, a number of mimeographed doctoral theses are available. Special bibliographies, two in English and one in German, on “Television as a technical aid in education and in educational and psychological research” have been published by Bierschenk (1969, 1971 b, 1971 c). A third report including all three, with an introduction in Swedish, appeared in Bierschenk (1971 d). A survey of literature on educational and psychological research into the techniques of audiovisual self-confrontation is given in Chapter 3.

The reactions of the teacher when the desired teaching behavior has been specified and accepted by him have been described and analysed in numerous studies. In contrast, the aim of the present study is to investigate the teacher's reactions when *no* specific norms have been externally and explicitly predetermined.

Like the actual teaching situations, the behavior of teachers and students can and has been studied in many different ways. Moreover, it has become increasingly obvious during the past few years that self-knowledge and self-understanding require quite different research methods than the study of curricula, material-method systems and teaching tasks.

Those working in educational-psychological research have long lacked the means of placing the complex process of teaching under experimental control. The technique of micro-lessons (a technique of reduction) has proved to be a very useful research method. The whole structure of micro-lessons can be manipulated in such a way that different problems can be answered by means of experimental designs. In addition, the CCTV system and the video-recorder make possible new approaches in research methodology for the study of interaction processes.

The main purpose of this study, "Self-confrontation via closed-circuit television", has been to study the reactions of student teachers placed under various experimental conditions when confronted with their own teaching performances, which have been registered by means of closed-circuit television and video-recording. An additional aim has been to study the "degree of objectivity" of the perceptions and evaluations of the student teachers, by examining their selection of information from the video-recorded teaching situations. The studies of effect concerned different forms of feedback, such as dyadic confrontations in the form of traditional tutorship and externally mediated self-confrontation processes via CCTV/VR.

To summarize, the goals have been to:

1. study systematically and under controlled conditions the way in which the student teacher perceives and evaluates the behavior of himself and his pupils in the context of micro-lessons.

As a result of the treatment of this problem, the individual's "self-concept" and "life-space cognition" have come to occupy a central position in the study.

2. examine systematically the dyadic confrontation process in a tutoring situation.

Since tutoring plays such a central role in teacher training, a detailed analysis of the pattern of "face-to-face" communication ought to be of considerable importance for educational research.

3. gain experience in the use of closed-circuit television, video-recorders and micro-lesson techniques as aids in research and as teaching methods.

The advantages of CCTV/VR and micro-lessons have been pointed out in various contexts and need not be further emphasized here. One disadvantage of CCTV, video-recording and micro-lesson techniques is that editing at certain stages can make the protocol material liable to subjectivity. Lighting and sound-recording can also cause difficulties.

For a more detailed discussion, see Chapter 2.

32. The experimental design of the study

Studies dealing with possible ways of using CCTV/VR techniques in educational psychology often seem to be characterized either by faulty designs or by inadequate models for the treatment of data.

Stickell's (1963, p. 46) investigation showed that of 250 data comparisons concerning "televised and face-to-face instruction", only 6 % were based on control group designs with "satisfactory control groups", while 50 % were not based on any experimental control. Stickell's examination showed (p. 48) that only 10 out of 250 comparisons led to interpretable results. Controlled experiments are, however, the "only way of verifying educational improvements" (Campbell & Stanley, 1963, p. 172).

The main problem in the present study has been:

What are the effects of traditional tutoring in the form of dyadic confrontation and/or externally mediated self-confrontation via CCTV/VR on the self-assessment (perception and evaluation) of student teachers?

Those participating in this study were 96 female student teachers, admitted in the autumn terms of 1967 and 1968 to the School of Education in Malmö for training as teachers in grades 4—6. They took part in the experiment at the beginning of their second term.

In order to achieve the maximum degree of control over possible interpretations of the results of the study, a factorial design was drawn up. The different factors of the design are:

Factor H: Traditional tutoring, where

h₁: tutoring

h₂: no tutoring

Factor T: Self-confrontation mediated externally via CCTV/VR, where

t₁: self-confrontation

t₂: no self-confrontation

Factor U: Micro-lessons (length 15 min.), where

u₁: micro-lesson 1

u₂: micro-lesson 2

In order to increase the precision of the design, two additional factors were included in the original design, i.e.

Precision factor V: Assessment and evaluation schedule F III

v₁, . . . , v₇₉ statements of which the measuring instrument consists.

Precision factor A: Aspects of the instrument. Each statement contains two different aspects, where

a₁: perception aspect

a₂: evaluation aspect.

The whole ANOVA model can be written as A, U, T, H, I (TH), V, where I denotes the factor representing the individuals. A summary of the experiment's analysis of variance design is presented in Table 75. In addition, a brief description is given below of factors H, T, and U (for a description of factors I, A, and V, see Chapters 4 and 8.2).

Table 75. The analysis of variance design of the experiment

Index	A	U	T	H	I	V
No. of levels	2	2	2	2	24	79
Size of population	2	2	2	2	∞	79

32.1 Description of factors H, T, and U

Factor H: Traditional tutoring in the form of dyadic confrontation was arranged so as to be similar to the tutoring student teachers receive during their teaching practice. Acting in the same way as during normal teaching practice, the tutor observed the student teachers during the experiment, i.e. made the notes considered necessary for the subsequent tutoring session.

The tutor was allowed the same length of time for discussing the lesson with the subject as is normal in ordinary teaching practice.

Factor T: Externally mediated self-confrontation via CCTV/VR here implies confrontation with one's own behavior in teaching situations which are registered via closed-circuit television and video-recording. The process involves external confrontation with one's own expressive behavior. These experiences of confrontation could be described as a de-automatization of the usual way of seeing one's self. This Factor T can thus be said to involve an external self-distancing in space and time.

In order to avoid or balance possible sources of error associated with technical problems such as camera angle or different methods of editing (spatial selection, temporal selection), two cameras were used, linked via a mixer, for registration of the behavior of the experimental subjects, while the behavior of the pupils was registered via a third static camera. To make it possible to examine the facial expressions of the subject, a close-up was registered by zooming in every third minute. The close-up lasted for 10 seconds. This measure was introduced as a result of a preliminary experiment (cf. Ch. 2) where the student teachers expressed a desire to see themselves in close-up. A more detailed description of the techniques used in the experiment for both recording and playback, together with the arrangement of the apparatus, are given in Chapter 6.

Factor U: Micro-lessons involve three different components, namely (1) pupils, (2) theme of the lesson to be taught and (3) length of lesson.

1. *Pupils* participating in micro-lessons should be representative for the level that the aspirant teacher is going to teach. The pupils (half-classes) that took part in 1969 and 1970 all came from the fourth grade of Munkhätteskolan in Malmö and were divided between the four experimental groups without regard to ability or social class.
2. *The theme of the lesson to be taught* was taken from the subject area of biology. Within this area the animals of northern Sweden were chosen: lemming (I), bear (II), wolf (III), lynx (IV), reindeer (V), golden eagle (VI), grouse (VII) and wolverine (VIII). (The Roman figures given in brackets refer to Table 8.) These subjects were chosen on the recommendation of lecturers in Methodology as being relatively simple. At this level there is no noticeable variation in the technique required either for teaching about the different species or for planning the teaching. An extra advantage in choosing biology was that all the student teachers had access to abundant and very similar concrete illustrative material.
3. *Length of lesson*, i.e. the video-recorded teaching time, was limited to 15 minutes. The student teachers were also allowed an additional 5 minutes warming-up time to get acquainted with the pupils. The short duration (15 min.) of the lesson forced the student teachers to keep to the task at hand, and imposed a natural restriction on too wide a variety of teaching activities.

Table 8 gives a summary of the experimental plan for the first micro-lesson. The second micro-lesson repeats exactly the same plan. A survey of the experimental design is presented in Figure 3.

The factorial design shown in Figure 3 is a more complex form of Campbell's and Stanley's design no. 6, "The Posttest-Only Control Group Design".

This design checks all the eight sources of error (see Campbell & Stanley, 1963, p. 178) that could invalidate the internal validity of the experiment. In addition, the experiment must have external validity if we are to make generalizations on the basis of the result for the population of student teachers in question (cf. Fig. 1).

A detailed discussion of the internal and external validity of the experiment, together with a description and discussion of some empirical results that illuminate certain aspects of the ecological validity of the experiment, has been presented in Chapter 11.

32.2 Data from attitude questionnaires

In order to find out to what extent the experiment was felt to be something exceptional compared to the usual teaching situation, three attitude questionnaires were constructed and administered to (1) the teaching staff at the School of Education, (2) the student teachers in the second term of their training as teachers in grades 4-6 who did not participate in the experiment, and (3) the student teachers in the second term of their training as teachers in grades 4-6 who did participate in the experiment. The answers to the separate questionnaires can be summarized as follows:

Teaching staff at the School of Education in Malmö

The teacher-trainers consider that:

1. the experiment described above is "very important" for teacher training
2. the participation of student teachers in such experiments is "very important" for teacher training and
3. the School of Education should in the future "to a very great extent" carry on research into the use of closed-circuit television in teacher training.

Despite the fact that the experiment was integrated into the teacher training schedule, some members of staff have complained that the experiment interfered with the normal course of their work.

Student teachers who did not participate in the experiment

The answers received from the student teachers who did not participate in the experiment can be summarized in the following way:

1. This group has a positive attitude to the CCTV experimental activity in both 1969 and 1970.

2. The student teachers wish to see their own lessons, registered via CCTV/VR, rather often—a positive attitude to the medium and the method concerned.
3. The student teachers are very hesitant as to whether they could accept losing some of their scheduled training. This reaction is completely in line with the current practice at the School of Education. Furthermore, the student teachers are hardly willing to accept a greater load of work.
4. The risk that the student teachers participating in the first phase of the experiment should have lost essential parts of the training was judged in 1969 to be minimal, while the student teachers participating in the second phase were thought to have lost essential lectures.

A possible explanation is that the greater difficulties involved in integrating phase II of the experiment into the schedule have caused a change of attitude among the student teachers. The variation in the number of lessons lost was namely greater in 1970 than in 1969. (The student teachers of 1970 who participated also considered their absence from their training to be more serious than those of 1969 had done.)

Student teachers who participated in the experiment

The answers received from the student teachers who participated in the experiment can be summed up as follows:

1. All student teachers found participation in the experiment enjoyable.
2. All student teachers considered it to be valuable experience to participate in such experiments during their teacher training.
3. All student teachers considered that continued research into the use of CCTV/VR techniques in teacher training should be carried out on a large scale at the School of Education.

To summarize, it can be said that the internal validity of this study is guaranteed by the factorial design presented above. That the experiment should have internal validity is a minimum requirement, without which one cannot even start to interpret an experimental study. It has also been pointed out that the possibility of generalizing the results of the experiment to the population in question is partly dependent on the ecological validity of the experiment.

In conclusion, the results of the attitude questionnaires show that the ecological validity of the experiment is reasonably good. The experiment has not been felt to interfere unduly with the teacher training program. Moreover, the general attitudes towards the experiment and self-confrontation via CCTV/VR have been positive.

32.3 Various sub-studies

In addition to the factorial design for examining the self-assessment of the student teachers, the arrangement of the study also permits investigation of the following problems:

1. Assessment of student teachers by educational experts

The micro-lessons of the student teachers have been assessed by four educational experts. These independent assessments have been examined for agreement between two of them. An analysis of variance model was then used to help find out whether, taking the "average assessment" of the experts as a criterion, participation in the experiment had resulted in any demonstrable effects upon the teaching behavior of the student teachers.

2. The self-assessment of the student teachers and the assessments of the educational experts

An important goal in teacher training is to develop the skill of the student teachers in interpreting educational processes "objectively", i.e. realistically. In order to study the "degree of objectivity" in the perception and evaluation of the student teachers, it is necessary to have an external criterion. In the experiment this external criterion consists of the "average assessment" of the educational experts. An examination of the "objectivity" of the student teachers' self-assessment, as defined below, also requires, however, that there should be a basis for assessment (e.g. video-recorded teaching situations) and a rating scale (categories) that are identical for both student teachers and experts. The operational definition of "objective" perceptions and evaluations used in this study is based on the experts' average assessment as criterion of the objectivity. As a measure of the deviation of the student teachers' self-assessment from this criterion, the differences have been calculated. This means that a large difference value indicates low objectivity and a small difference value indicates high objectivity in the student teachers' self-assessment. (The expression "deviations in objectivity" refers to variations in the calculated difference values.) The way in which the difference between the experts' average assessment and the student teachers' self-assessment varies as a result of the experimental treatment has been studied by means of an ANOVA (Analysis of Variance).

Furthermore, the observations of the educational experts and the student teachers have been studied with a view to finding possible similarities in struc-

ture. The possible occurrence of an overlapping structure between the self-assessment of the student teachers and the average assessment of the experts has been investigated by canonical correlation analyses.

3. The influence of student teachers' predispositions and personalities on their own perception and evaluation of teaching process mediated by CCTV/VR techniques

In order to be able to study a possible connection between the student teachers' special perception and evaluation tendencies on the one hand, and aspects of their personality on the other, a battery of group test has been administered, containing different personality tests, cognitive tests and attitude tests. For a detailed presentation of the test battery, see Chapter 8.2.6.

4. Follow-up studies

The student teachers were asked to assess the video-recorded micro-lessons again, first six weeks and then four terms after the experiment had been concluded. The purpose of this follow-up was to examine to what extent the teacher training had had any effect on their perception and evaluation of the micro-lessons video-recorded during their second term at the School of Education. An analysis of the special studies in points 3 and 4 has not been included, however, in the present phase of reporting.

33. Assessment and evaluation schedule F III: Validity and reliability

The great majority of studies using CCTV/VR techniques as an instrument for research and training have relied on well-known test methods. If one wants to find the answer to a particular problem, however, it is not usually possible to apply old tests to the new problem.

The development of the measuring instrument "assessment and evaluation schedule F III" (cf. Bierschenk, 1972 f and part 1) started with a preliminary experiment (spring term 1968). The construction of the measuring instrument was based on the following question: What do the student teachers really tell us when they are confronted with their own teaching performances by means of CCTV/VR? Thus the measuring instrument was developed from scratch.

The comments made by the student teachers in the experiment during the process of self-confrontation were recorded on tape and then subjected to content-analytical treatment. This treatment could be carried out with a fairly satisfactory coder agreement, both in deciding the coding units ("information units") in the comments of the student teachers and in coding the information units in accordance with a system of categories (cf. Bierschenk, 1972 f). Thus this treatment resulted in an acceptable coder agreement in the categorization of the student teachers' spontaneous and simultaneous comments during the playback of the video-recorded micro-lessons. The separate categories have then been re-formulated into the statements which make up the assessment and evaluation schedule F III. The results presented in this report are based almost exclusively on this instrument. The statements included in the measuring instrument thereby reflect the problem areas on which the student teachers themselves have spontaneously focused attention. The content analysis only makes it possible to judge the content validity of the instrument, but this is a very essential and desirable form of validity. It is often missing from psychological tests on personality, or is only said to exist on flimsy grounds. It is logically impossible to define "the whole process of perception" and since we cannot specify the process of perception completely, we cannot know whether or not we have a measuring instrument of "true validity". We must, therefore, be content to estimate the validity indirectly.

The problem areas which emerged from the student teachers' simultaneous comments during the process of self-confrontation have been categorized according to the following six dimensions constructed a priori:

1. ego-ego relation
2. ego-pupil relation
3. ego-NPO relation
(i.e. relation between
ego and non-personal
objects)
4. pupil-ego relation
5. pupil-pupil relation
6. pupil-NPO relation

These dimensions are defined in the assessment and evaluation schedule F III by a total of 79 statements. The instrument has been constructed to quantify the individual's ability to (a) assimilate information (perceive) (b) digest information (evaluate) and (c) use this information (modify).

If one wishes to prove the existence of symmetrical relations or to be able to predict a particular type of behavior, the measuring instrument must give evidence of being a reliable measure. A certain amount of information about the reliability of the self-assessment of the student teachers can be gained from an examination of the communality of the variables (h_i^2), i.e. the squared multiple correlation, R_{ik} , can be used as an estimation of the lower limit of reliability of a particular item. If the decision as to whether a certain item has unsatisfactory common variance is based on the criterion $h_i^2 < .30$, then four statements within the area of perception (aspect a_1) and three statements within the area of evaluation (aspect a_2) fail to fulfil this limit value.

Since a separate ANOVA has been carried out for each individual subject-object relationship, summed up over the entire variable domain, an attempt was also made to calculate the reliability of the individual variable domain by means of "Cronbach's alpha". The result indicated, however, that in this case this coefficient is not a suitable measure of reliability. Therefore, the following have been presented as comparison criteria:

- (1) the squared average multiple correlation
- (2) the average communality and
- (3) an estimation of homogeneity according to Spearman—Brown's "prophecy" formula.

Reliability can be defined and estimated in many different ways and for that reason estimation based on a single method can easily lead to over-hasty conclusions.

The examination of reliability has established that, as a whole, the reliability of the student teachers' self-assessment is satisfactory. The reliability of the

individual statements in the educational experts' assessments has been estimated by means of the intra-class correlation coefficient, r_{21} , i.e. for two experts and one teaching occasion. The reliability of the summation variables was also calculated with "Cronbach's alpha".

A closer examination of the reasons for the unreliability of individual items showed that the decision as to the reliability of a particular item could not be based solely on r_{21} . Both m and s and the raw data indicated that for some items there was no response variation at all. In such cases a reliability measure based on variance cannot be used to indicate the agreement in the experts' estimations. For the same reason, "Cronbach's alpha" does not appear to be altogether suitable as a measure of reliability, even if satisfactory reliability coefficients could be demonstrated for certain variable areas (because of larger distribution values).

In summary, it may be said that the low reliability values obtained from the different reliability measurements based on variance have often been caused partly by a lack of variance within certain variable areas, and partly by the low item covariance. The relatively extensive examination of reliability has shown that the assessment and evaluation schedule F III can be regarded as being essentially a reliable measuring instrument.

34. Analyses of results

Data for the complete design exist for the assessment and evaluation schedule F III. This schedule has been made the primary subject for analysis and these observation data have been analysed for both level and structure. The individual analysis programs have been described in Chapter 10. The summary and discussion of results presented in this report concern above all:

1. ANOVA treatment of the student teachers' self-assessment (part 2),
2. ANOVA treatment of the assessment by educational experts (part 3), and
3. ANOVA treatment plus a canonical correlation analysis of the student teachers' self-assessment and the educational experts' assessments (part 4).

Problems of research method were discussed in comparative detail in connection with the individual result analyses. Completed result analyses that have not yet been reported include a number of factor analyses and simultaneous comments which have been coded.

If any form of inference statistics is used in the analysis of behavioral observation data, then (1) the prerequisites demanded by a particular statistical mathematic model should be fulfilled, and (2) the precision and power of the statistical tests used should have been explicitly determined.

As omnibus tests, significant F tests are very useful indicators of systematic differences among cell means, but only a careful examination of detail will make it possible to interpret the experimental results thoroughly. If, in addition, the design is rather complex, a large number of F tests are needed, and that in its turn increases the probability of a certain number of tests resulting in random significances. For this reason, one should avoid attaching too much importance to isolated results. The guideline followed in the evaluation of the analyses of results has therefore been the interpretability of the patterns in the F tests. In order to obtain additional and more objective indicators as to whether there is any point in a more thorough interpretation of the main and interaction effects respectively or in carrying out contrast analyses and in commenting on simple effects, the precision and power of the significant F tests have also been calculated.

Thus the individual ANOVA results have been evaluated step-wise. First (1), the interpretability of the patterns in the F tests was examined. Then the precision and power were estimated in order to decide (2.1) the size of the effects and (2.2) the probability of discovering an effect of a particular size. Contrast analyses (3) were not carried out until this point. Only effects at the level of $\alpha = .001$ which show a probability of at least .70 of discovering a particular size of effect have been interpreted, however. Values lower than this are of little use as evidence or as a basis for interpretation.

34.1. Analysis of levels

34.1.1 Step 1: Patterns in the F tests

The interpretation of the first step in the individual analyses of results (parts 2, 3 and 4) shows that the F tests in all three parts of the analysis have led to interpretable patterns. A summary of the patterns in the F tests, referring to the respective part of the analysis, is given in Table 76. Owing to the construction of the F tests as omnibus tests, the interpretation cannot go further at this stage of the analysis than to establish that there is a systematic pattern in the F tests, that in addition permits an interpretation that is meaningful from the point of view of educational psychology.

The pattern in the F tests for the self-assessment of the student teachers shows no significant main effect either in Factor T (self-confrontation via CCTV/VR) or in Factor H (dyadic confrontation in the form of traditional tutoring). In addition, H_0 is accepted for the factor combination TH.

The demonstrable interaction effects partly imply, however, that externally mediated self-confrontation via CCTV/VR leads to reactions that are influenced by the predispositions and predictions of the student teachers and that these reactions are modified by the information that the student teachers have received. It is also worth noting that the student teachers, regardless of the type of influence, seem to modify their perceptions and evaluations from lesson to lesson as far as their own behavior towards an object is concerned (ego dimension). There is no such modification, however, in the subject-object relationships where the student teachers must relate the actions of other persons to their own person (pupil dimension).

With regard to the summation variables, the pattern in the F tests for the average assessments of the experts shows with respect to the TH and ATH interaction effects a greater degree of homogeneity and more significances than the pattern in the F tests for the self-assessment of the subjects. Moreover, the average assessment of the experts in the pupil dimension (pupil-ego, pupil-pupil relationships) has led to significant main effects in Factor H.

An examination of the influence of traditional tutoring with regard to the interaction effects (Factors U, H) shows that these interaction effects (UH, AH) largely apply to the pupil dimension. One possible explanation is that as a result of the tutor's influence the climate of behavior both between student teacher and pupils and among the pupils themselves have changed. But these effects can also simply be a consequence of the presence of the tutor in the classroom, which can have had a subduing effect on the pupils. In this context, it is not possible to decide which of these explanations is the most probable.

As far as the interaction effects involving Factor U are concerned, the analysis of the self-assessment of the student teachers shows that these effects

Table 76. Summary of the significant F tests for the summation variables in the respective sub-analyses

Source	1			2			3		
	Ego-ego relation			Ego-pupil relation			Ego-NPO relation		
	1	2	3	1	2	3	1	2	3
T									
H									
TH	**	*		*	**				
U	**		**	**	*		**	**	
UT									
UH									
UTH									
A	**	**					**	**	
AT			*				*		**
AH									
ATH	**	*		*	**	**			
AU	**						*		*
AUT							**		**
AUH							*		
AUTH									
4			5			6			
Pupil-ego relation			Pupil-pupil relation			Pupil-NPO relation			
1	2	3	1	2	3	1	2	3	
T									
H									
TH	*	*	*				**		
U	**	**					**		
UT	*								
UH	**			*	**		*		
UTH									
A	**	**	**	**	**		**	**	
AT									
AH							*		
ATH	*	**	**	*	*		**	**	**
AU	**		*						
AUT				**					
AUH							**		**
AUTH									

1: Analyses of student teachers' self-assessment

2: Analyses of educational experts' assessments

3: Analyses of student teachers' self-assessment and educational experts' assessments

** $F_{.99} (1, 92) = 7.08$

* $F_{.95} (1, 92) = 4.00$

are also mainly restricted to the pupil dimension. Within the ego-NPO relationship both student teachers and experts appear to have observed changes between microlessons 1 and 2.

The significant interaction effects for the factor combinations TH and ATH imply that the combination of T and H produces demonstrable differences between the groups. No effect can be seen in the UTH interaction, however.

The pattern in the F tests has been examined for variations in the objectivity of the student teachers' self-assessment. The interaction effects involving CCTV/VR imply that the processes of self-confrontation have resulted in demonstrable variations in the differences between the student teachers' self-assessment and the experts' average assessment. These effects indicate that there is a variation in the objectivity of the student teachers' assessment of the relations in which ego is the subject. Demonstrable variations in the differences between the assessments of the student teachers given traditional tutoring in the form of dyadic confrontation and those not given this treatment, however, relate to the pupil dimension.

Moreover, the effects in Factor U show that, regardless of the factors T and H, the objectivity of the observations of the student teachers have been influenced by the micro-lessons.

Thus, on the question of the ego-ego relationship (the student teacher's own person) and the behavior of the pupils towards the student teacher (e.g. the extent to which the pupils are "socially provocative"), the difference between the assessment of the student teachers and the experts average assessment shows noticeable variations from micro-lesson 1 to micro-lesson 2. It should be noted that both effects relate to the variable domains where ego is the object.

For Factor A the F tests show a very homogeneous pattern for the observation data of both student teachers and experts (cf. Table 76, sub-analysis one and two). This means that, regardless of the experimental conditions and specific statements, the perception (a_1) and the evaluation (a_2) are different. The homogeneous pattern that appears in all the variable domains is an expected outcome. This factor was included in the analysis as a precision factor and not because of any wish directly to compare perception and evaluation.

In part three of the analysis, Factor A shows demonstrable effects within the ego-pupil relation and within the pupil-ego relation. This result implies that there are differences in objectivity between the perception and evaluation of the student teachers (regardless of the experimental conditions or lesson concerned) within these sectors.

It is not possible to make any deeper interpretation in this step of the analysis. Only an examination of the significant F tests by means of contrast analyses can provide information as to what has caused these significances.

34.1.2 Step 2 of analysis: Precision and power in the F tests

Summaries and discussions of results have been reported in more detail in parts 2, 3, and 4. In these discussions, however, the effects have only been taken into consideration when precision and power estimations have indicated that there is a conclusive basis for interpretation. The predictor variance ($\hat{\omega}^2$) in the significant effects proved to be relatively low in all three parts of the analysis. But since the numerical size of $\hat{\omega}^2$ is dependent on how many sources of variation are included in an ANOVA, a comparison criterion is needed if one is to be able to decide on an objective basis whether the $\hat{\omega}^2$ values presented here are indications of unimportant correlations or perhaps of important empirical results. J. Cohen's (1969) "effect size index" (f), which gives a certain effect size (ES) when all other effects in the analysis have been held constant, has been used for this purpose. Cohen, J. (1969, p. 278) denotes a small effect with $f = .10$. A medium effect corresponds to $f = .25$ and a large effect to $f = .40$.

If f is calculated, Cohen's tables (pp. 282—347) can also be used to decide the power of the F tests. The tables must be used with a certain amount of care, however, since in factorial designs df in the denominator no longer agrees with the n values stated in the appropriate table. In factorial designs there is usually a lower number of df in the denominator than the number stated for the table values. As a result, above all as far as the main effects are concerned, the power that can be read in the tables is an over-estimation of the probability of the effect in question. Main effects with an $f < .35$ constitute an uncertain basis for interpretation (for a more detailed discussion, see Ch. 15.2).

A summary of the probability values (g), referring to the respective part of the analysis concerned, is given in Table 77.

Information on the precision values of the effects ($\hat{\omega}^2, f$) can be obtained from parts 2, 3, and 4. With these estimations of power it becomes possible not only to state that an effect of a particular size does exist, but also to state the degree of probability that this effect really is of the stated size. Table 77 shows how the step-wise procedure has finally produced only a few effects within each separate sub-analysis that are suitable for more detailed consideration and interpretation. In the light of this result, there has been no reason to undertake a more thorough interpretation of the contrast analyses.

34.1.3 Step 3: Post-hoc comparisons

Since it can be difficult to define what is to be regarded as really valuable information, the controls described above were applied in order to decide whether or not a more detailed examination and discussion of the results of the experiment would be worthwhile.

As was pointed out (Chs. 3 and 11), there have unfortunately been far too many cases in which no attempt has been made to state to what extent

Table 77. Summary of power values for the significant F tests ($\alpha = .01$) in the three parts of the analysis (summation variable)

Source	1			2		
	Ego-ego relation			Ego-pupil relation		
	1	2	3	1	2	3
T						
H						
TH				.62		
U				> .99		
UT					.78	
UH					.78	
UTH						.62
A				> .99		
AT					.72	
AH						> .99
ATH				.66		
AU					.91	
AUT						.93
AUH						
AUTH						
<hr/>						
Source	3			4		
	Ego-NPO relation			Pupil-ego relation		
	1	2	3	1	2	3
T						
H						
TH						.72
U				.83		
UT					.67	
UH						.83
UTH						.61
A				> .99		
AT					.52	
AH						.67
ATH						
AU						.99
AUT						
AUH						
AUTH						

the requirements for a given method of analysis have been complied with, or to decide the precision and power of the tests in question. Only after such estimations have been made, however, is it possible to judge the usefulness of the significant F values for a more detailed study of the relationship between the simple effects. An evaluation of the experiment's data with the F statistics involves testing the null hypothesis. Rejection of the null hypothesis implies that the set of data in question contains systematic effects. An F test does not indicate the direction of the effects, however, nor does it state the precision of the measurement or the probability of an effect being of a certain size. Seen in the light of the power estimations to be found in Table 77, the self-assessment of the student teachers could not on the whole be used for a detailed analysis of the contrasts, and for this reason the contrast data were presented as an appendix (Bierschenk, 1972 e). Compared to the self-assessment of the student teachers, the average assessment of the experts at least with regard to the TH and ATH interaction effects has resulted in a more uniform pattern in the F tests, in greater precision, and in a higher degree of probability for the proven effects. This type of result was expected, however. A large proportion of the variation in the student teachers' self-assessment can probably be traced to differences between the individuals that existed prior to the experiment. Since the analysis of the student teachers' observation data is based on $n=96$, while the analysis of the average assessment is based on $k=2$, the standard deviation of the means is smaller in the experts' observations. This in its turn means that the differences between the cell means need not be as large as for the student teachers in order to produce demonstrable effects.

Since there are always deficiencies which could have led to small effects and low probabilities, an account of the contrast analyses may be of interest for further research work. Any reader who is also interested in result analyses 2 and 3 can refer to Bierschenk, 1972 e. Relatively few significances have shown a satisfactory power in the effects and these are therefore discussed in the chapter "Final discussion" (see parts 2, 3, and 4).

In order to obtain a more surveyable perspective of the tendencies that seem to appear in the separate analyses, a summary of the main effects is given in Table 78.

Table 78 shows how the average assessment of the experts alone resulted in demonstrable main effects in Factor H of the experiment. The mean values indicate a more positive assessment of the group with dyadic confrontation.

The self-assessment of the student teachers has for Factor U resulted in three demonstrable effects, all of which involve the *ego* dimension. According to the average assessment of the experts, however, the demonstrable effects primarily involve the pupil dimension. But both student teachers and experts have reported changed values in variable domain 3, i.e. in the re-

Table 77. (Cont.)

Source	5 Pupil-pupil relation			6 Pupil-NPO relation		
	1	2	3	1	2	3
T						
H			.78			
TH						.52
U			.46			
UT						
UH				.84		
UTH						
A	>.99	>.99				
AT						
AH	.52	.89				
ATH					.76	.73
AU						
AUT	.55					
AUH					.69	
AUTH						.85

1: Analyses of student teachers' self-assessment

2: Analyses of educational experts' assessments

3: Analyses of student teachers' self-assessment and educational experts' assessments

lation between the student teachers and the aspects concerning teaching method. In both cases, the change was positive. In this factor, significant variations in the objectivity of the student teachers' self-assessment can be demonstrated in two cases. In the ego-ego relationship the difference is greatest in connection with lesson 1 and diminishes strongly in connection with lesson 2. The same tendency can be observed in variable domain 4.

Factor A represents two aspects of the measuring instrument, namely perception (a_1) and evaluation (a_2). Differences between perception and evaluation appear in each variable domain for both the self-assessment of the student teachers and the average assessment of the experts. The tendency in both sets of data is the same. The mean values seem to indicate a tendency for the perception of the experts to be more positive than that of the student teachers.

Variations in the differences between the self-assessment of the student teachers and the objectivity criterion are significant only for variable domains 2 and 4. There is a very slight difference between the objectivity criterion and the student teachers' perception regarding the pupil-ego relation (4). No socially provocative behavior on the part of the pupils seems to have occurred. In their evaluation, however, the deviation is comparatively great. The student

teachers evaluate possible behavior of this kind as being rather distressing, while the average assessment of the experts is that it is relatively easy to deal with.

Table 78. A summary of the mean values of the significant main effects for the partial analyses 1, 2, and 3

Variable domain	Student teachers' self-assessment (1)		Experts' average assessment (2)		Student teachers' self-assessment in relation to experts' average assessment (3)	
	Factor H		Factor H		Factor H	
	h_1	h_2	h_1	h_2	h_1	h_2
4	—	—	(5.21	5.14)	—	—
5	—	—	(4.85	4.61)	—	—
	Factor U		Factor U		Factor U	
	u_1	u_2	u_1	u_2	u_1	u_2
1	4.60	4.80	—	—	(.43	.28)
2	(4.64	4.74)	—	—	—	—
3	(4.90	5.04)	(4.95	5.01)	—	—
4	(4.93	5.00)	(5.21	5.15)	(.27	.14)
5	—	—	(4.80	4.66)	—	—
6	—	—	—	—	—	—
	Factor A		Factor A		Factor A	
	a_1	a_2	a_1	a_2	a_1	a_2
1	4.90	4.50	5.22	4.90	—	—
2	(4.77	4.62)	5.04	4.59	.28	.03
3	4.66	5.28	4.68	5.28	—	—
4	6.12	3.82	6.15	4.20	(.03	.38)
5	4.88	4.17	5.11	4.33	—	—
6	5.27	5.62	(4.85	5.03)	—	—

(-) Uncertain basis for interpretation.

34.1.4 Implications of ANOVA results

The use of different analysis techniques and the stepwise approach used in reporting the results aim at making the evaluation more critical and thereby more objective.

We hope that the result analyses can serve as examples of the application of principles of research method and of how one can explicitly prove if and to what extent the assumptions of the statistical-mathematical models are fulfilled. In this way the inferences become meaningful.

In addition, it is hoped that the detailed description of the experimental conditions (Chs. 4 and 5) will provide:

1. increased knowledge of possible ways of using CCTV/VR as a research and training instrument,
2. increased possibilities of repeating behavioral experiments or at least
3. increased opportunities for comparing individual research results.

Finally, keeping in mind Stickell's (1963) examination of research reports concerning "televised and face-to-face instruction", the results can be seen as

4. a contribution to improving the quality of research results dealing with the use of CCTV/VR in educational contexts.

From an empirical point of view this experiment has produced results which cannot be made the basis of dichotomous decisions, i.e. either-or decisions. If the experimental results are interpreted purely pragmatically or from the point of view of economy, it might seem reasonable simply to recommend the cheapest alternative, i.e. the student teachers seem to need no tutorship in the form of dyadic confrontation and/or externally mediated self-confrontation via CCTV/VR, since the experimental conditions have not led to any demonstrable main effects.

Speaking against such a decision, however, are certain alternative explanations of the null hypothesis and the fact that one cannot draw such far-reaching conclusions from a single experimental result. Another important argument against such a decision is the consideration of the possible long-term effects and this aspect will be examined more closely in the experiment's follow-up studies.

There can be many reasons for accepting the null hypothesis for a particular factor or factor combination, and therefore it is difficult to say anything definite when the null hypothesis is accepted. Despite this, a section entitled "Final discussion" has been included for each part of the analysis, presenting some more or less speculative considerations that could be regarded as possible implications of the results. In the following, we try to summarize point for point. The first figure states the order of sequence. The figure following the colon denotes the part of this study in which the result has been discussed in detail. Thus in the case of the self-assessment of the student teachers, for example, (1:2) states the results of both externally mediated self-confrontation via CCTV/VR (Factor T) and dyadic confrontation in the form of traditional tutoring (Factor H), (2:2) states the results of the teaching situations (Factor U), (3:2) states the results of the aspects in the assessment and evaluation schedule (Factor A), (4:2) states the results of the conclusive higher order interaction effects. After the results given in each sub-section, possible implications are presented.

The educational and psychological implications that have already been presented in connection with the individual experiment results have been divided between the following two categories: (1) Implications based on the

results of the experiment and (2) implications based on more speculative considerations.

The first category covers implications that either (1) are of a descriptive nature, i.e. establishing facts, or (2) are based on experimental data which provide a conclusive basis for interpretation. The second category covers implications that either (1) are based on experimental data which do not provide a conclusive basis for interpretation or (2) are of a purely speculative nature.

The borderline between the two categories, however, can doubtless be challenged in many cases. The main reason for making this division was to make clear which implications can be said to be based on the *conclusive* effects of the experiment. Another motive was that the large number of implications presented could give an undesirable impression of indecision. Finally, some results and implications have been further clarified by short comments.

1:2 Results: Student teachers' self-assessment

Neither externally mediated self-confrontation via CCTV/VR nor dyadic confrontation in the form of traditional tutoring (Factor H) have led to significant effects.

Implication, based on these results

The experimental conditions produce no effect, i.e. have not led to any difference in the ability of the student teachers to discriminate.

Implications, based on more speculative considerations

1. The experimental conditions lasted for too short a time for the various influences to achieve observable effects.
2. Self-confrontation requires systematic training in receiving and adapting first-hand information, i.e. "self"-information not mediated verbally.
3. Self-confrontation entails a temporary de-organization or de-automation, the first phase of which produces in many people feelings of surprise, fear, shock and/or the adoption of defensive attitudes.
4. The tutor has not succeeded in influencing the student teachers to any degree since they have not yet developed suitable test criteria, i.e. educational-psychological norms.
5. Tutor and student teacher avoid a relevant critical analysis by, e.g. using words such as "a verbal portrait of an individual" (Stoller, 1970, p. 11) in order to avoid having to make a critical examination of the student teacher's own behavior.

1:3 Results: Educational experts' assessments

The null hypothesis is accepted for externally mediated self-confrontation via CCTV/VR (Factor T), but is rejected for dyadic confrontation in the form of traditional tutorship (Factor H). Significant effects have been demonstrated within pupil-ego relations (4) and pupil-pupil relations (5).

The estimation of precision and power indicates, however, that these effects can hardly be regarded as an acceptable basis for interpretation.

Keeping in mind the far from conclusive effects, a few possible but rather more hypothetical interpretations are presented below.

Implications, based on more speculative considerations

1. The tutor has influenced the student teachers in such a way that their behavior has become more positive. Thus the tutor has successfully mediated both his teaching strategy and some concrete suggestions for action and the student teachers have succeeded in modifying their own behavior on the basis of the dyadic confrontation. In addition, the result seems to be in agreement with the tutor's intention, namely to focus the tutorship upon problems of pupil activation.
2. A prerequisite of traditional tutorship is the presence of the tutor during the actual teaching process, and there is therefore a possibility that these effects have arisen as a result of the subduing effect that the teacher's presence has had on the pupils' activity.

The second alternative (2) seems the most probable, since none of the relations where the student teacher is the subject have led to demonstrable effects.

1:4 Results: Student teachers' self-assessment and educational experts' assessments

Neither externally mediated self-confrontation via CCTV/VR (Factor T) nor dyadic confrontation in the form of traditional tutorship (Factor H) have led to significant effects.

Implication based on these results

Since no deviations in the differences between the student teachers' self-assessment and the average assessment of the experts can be demonstrated, the objectivity, e.g. the "gap" between the objectivity criterion and the student teachers' self-assessment has not been influenced by the experimental treatment. In the context of the definition of objectivity applied in this study, the deviations in objectivity in the observations of the student teachers do not vary as a consequence of the experimental treatment.

2:2 *Results: Student teachers' self-assessment*

The teaching situations, i.e. micro-lessons 1 and 2 (Factor U), have resulted in significant effects concerning the ego dimension: ego-ego relation (1), ego-pupil relation (2), ego-NPO relation (3). For the pupil dimension (variable domains 4—6), the null hypothesis is accepted. The estimations of precision and power indicate that the effect within the ego-ego relation may be looked upon as a conclusive basis for interpretation.

Implications, based on these results

1. Taking micro-lesson 1 as a starting-point, the student teachers seem to be able to predict their own behavior and test these predictions during micro-lesson 2, and subsequently to modify the structure of perception and evaluation, or the concrete behavior.

As far as the student teachers' perception and evaluation of the pupil dimensions is concerned, the observations indicate no changes.

Implications, based on more speculative considerations

1. The pupils' behavior has not changed (pupil as subject).
2. The student teachers lack criteria for assessing the behavior of the pupils.
3. The student teachers have been primarily occupied with their own person and have therefore not had time to study the behavior of the pupils in any detail.

2:3 *Results: Educational experts' assessments.*

For the teaching occasions (Factor U) the null hypothesis is rejected within the ego-NPO relation (3), pupil-ego relation (4) and pupil-pupil relation (5). The mean values indicate a positive change with regard to the ego-NPO relation, while the change for the other relations (4, 5) is negative. Estimations of precision and power, with the possible exception of the pupil-ego relation (4), indicate an inconclusive basis for interpretation, and for that reason the interpretations suggested below should be regarded as being hypothetical.

Implications, based on more speculative considerations

1. The examination of effects implies that the behavior of the pupils becomes more disturbing during lesson 2, despite the fact that the student teacher's teaching technique improves. This could mean that there is a "relaxation effect". It is possible that the first micro-lesson

was rather tense, since neither the pupils nor the student teacher were accustomed to the situation, while the second lesson could have been felt by both student teacher and pupils to be a more "everyday" situation, i.e. the behavior of the pupils has been more normal, in other words "more disturbing".

2. The small change in quality of the student teacher's teaching method has little significance for the teacher-pupil relation or, in other words, for the climate in the classroom.

2:4 *Results: Self-assessment of student teachers and average assessment of educational experts*

For the teaching occasions (Factor U) as cause of variation, the null hypothesis was rejected for the ego-ego relation and the pupil-ego relation. The mean values of the cells indicate variations in the objectivity of the student teachers' self-assessment that were greatest in micro-lesson 1, but diminished substantially in micro-lesson 2. An examination of the precision and power in this result implies that there are no conclusive empirical grounds for interpretation of the effects.

Implications, based on more speculative considerations

1. The diminished deviation from the objectivity criteria of the student teachers' self-assessment from micro-lesson 1 to micro-lesson 2 implies a more realistic assessment of their performances.
2. From the point of view of the student teachers, it seems obvious that progress has been made on the second teaching occasion. According to the educational experts, however, this is not the case.

3:2 *Results: Student teachers' self-assessment*

The perception and evaluation of the student teachers are included as Factor A (aspect) in the analysis of variance. The null hypothesis is rejected for this factor within all six subject-object relations. The assessments of precision and power show very high values, with the exception of the ego-pupil relation. This means that the effects can be regarded as providing a conclusive basis for interpretation.

Implications, based on these results

1. Irrespective of the experimental conditions, the student teachers' perception differs from their evaluation. Moreover, Table 78 shows that their perception has resulted in positive scores.
2. Irrespective of the experimental conditions, the student teachers have evaluated these behavioral aspects as being essential and undisturbing, with the exception of the pupil-ego relation where the student teach-

ers indicate that "socially provocative behavior" (if it had occurred) would have been considered relatively distressing.

3. A comparison of the student teachers evaluation, which relate to their perception of the behavioral aspects constituting variable domains 4 and 5, imply that the student teachers appear to have a high level of tolerance when it comes to the behavior of the pupils towards each other (variable domain 5), even if it is felt to be comparatively undisciplined, while direct action on the part of the pupils against the student teacher (with a conscious or unconscious element of provocation in it) is felt to be distressing.

3:3 Results: Educational experts' assessments

The experts' perception and evaluation (Factor A) show significant effects in all six subject-object relations. With the exception of the pupil-NPO relation, the estimations of precision and power have led to very high values. In this context, the effects can safely be regarded as a satisfactory basis for interpretation.

Implications, based on these results

1. Irrespective of the experimental conditions, the experts' perception differs from their evaluation. The experts' perception has in each variable domain resulted in positive scores.
2. Irrespective of the experimental conditions, the experts have in each case evaluated these behavioral aspects of the situation in question as being both essential and undisturbing (cf. 3:2).

3:4 Results: Student teachers' self-assessment and educational experts' assessments

With regard to the objectivity of the perception and evaluation of the student teachers (Factor A), the null hypothesis is rejected within ego-pupil (2) and pupil-ego (4) relations. Table 78 shows that the student teachers' perception of the ego-pupil relation differs negatively from that of the experts, while there is only a slight positive deviation in the evaluation. Within the pupil-ego relation, the situation is the exact reverse. The estimations of precision and power show, however, that only the effect within the ego-pupil relation provides a conclusive basis for interpretation.

Implication, based on these results

1. Irrespective of the experimental conditions, the deviation in the objectivity of the student teachers' perception of their own actions towards the pupils (2) is negative.

Implication, based on more speculative considerations

1. The effect size for the pupil-ego relation can hardly be said to indicate a conclusive basis for interpretation. While the student teachers evaluate socially provocative behavior as rather distressing, the experts evaluate it as being comparatively undisturbing. In his role as a leader, the teacher has a decisive influence on the social-psychological structure. His interpretation of what is socially provocative behavior should, therefore, be highlighted to a greater extent in educational contexts.

Externally mediated self-confrontation via CCTV/VR (Factor T) and dyadic confrontation in the form of traditional tutorship (Factor H), i.e. the factor combination TH, have only led to demonstrable effects in connection with the experts' average assessment. For the TH interaction the null hypothesis is rejected within the ego-ego relation (1), the ego-pupil relation (2), and the pupil-NPO relation (6). A possible explanation of the effects in the TH interaction is that the experiment was carried out in two phases, with a year's interval, which could have caused a change in the experts' perception and evaluation structure. The estimations of precision and power imply, however, that these effects should not be made the basis of any interpretation.

The experimental factors T and H plus the factor combination TH have also been examined for signs of interaction with Factor A (aspect) and Factor U (teaching occasion). A short summary is given below of results that fulfil our criteria for interpretation.

4:2 Results: Student teachers' self-assessment

Only a few of the interaction effects fulfil the requirements. The following interactions have been examined in more detail: (1) AUT within the ego-NPO relation, (2) ATH within the pupil-NPO relation and (3) AUH within the pupil-NPO relation.

Externally mediated self-confrontation via CCTV/VR (Factor T) in combination with aspect (Factor A) and/or teaching occasion (Factor U) led to changes within the ego dimension (variable domains 1—3). On the other hand, dyadic confrontation in the form of traditional tutoring (Factor H) in combination with Factor A and/or U resulted in changes within the pupil dimension (variable domains 4—6).

1. When considering the AUT interaction the perception of student teachers receiving only externally mediated self-confrontation via CCTV/VR shows no change. There is, however, a tendency for the evaluation to become more positive.
2. The perception of the student teachers who were not given this treatment showed a positive change, while the group's evaluation appears to be relatively unchanged.

3. When considering the ATH and AUH interaction the perception of the student teachers who received only dyadic confrontation is demonstrably more positive than the perception of those not given this treatment. However, through dyadic confrontation the student teachers' perception became demonstrably more negative.
4. The perception of the student teachers who were not given this treatment showed a positive change.
5. The evaluation reflects the tendency for the evaluation of the student teachers receiving traditional tutoring to change positively, while the evaluation of those not influenced in this way changes negatively. No significant differences between the simple effects for points 1, 2, 4 and 5 were demonstrable.

Implications, based on these results

1. It seems reasonable to assume that above all the student teachers receiving only externally mediated self-confrontation via CCTV/VR should have shown a change in perception, while the perception of those not receiving this treatment should have been more constant. An unexpected tendency is for the evaluation to change as a result of the self-confrontation via CCTV/VR, despite the fact that the evaluation structures are possibly more difficult to influence. This effect must naturally be looked upon as a relatively isolated result, but nevertheless it seems to point in the same direction as the result obtained by Perlmutter et al. (1967, pp. 900—905), namely that the subjects incline to change their structure of evaluation first.
2. The fact that the perception of the student teachers receiving only traditional tutoring (dyadic confrontation) changes negatively can possibly depend upon the fact that the tutor has pointed out to them certain concrete aspects of their behavior while at the same time verifying their evaluations.

4:3 *Results: Educational experts' assessments*

Several of the interaction effects fulfil our requirements for interpretation. However, Factor T in combination with Factor A and/or U has neither within the ego dimension nor within the pupil dimension resulted in demonstrable interaction effects. Factor H in combination with Factor A and/or U has, on the other hand, only led to significant effects within the pupil dimension. With respect to our criteria for interpretation, the following interaction effects have been examined in more detail: (1) UH within the pupil-ego relation and (2) AH within the pupil-pupil relation.

Furthermore, the ATH interactions in the variable domains 1, 2, and 6 show medium and large effect sizes. But since the educational experts'

assessment of the protocol material was carried out in two phases (first phase 1969, second phase 1970), it cannot be precluded that part of the variance may be traceable to a change in the experts' structure of perception and evaluation. With regard to the UH and AH interactions, the results can be summarized as follows:

1. The educational experts assess the teaching performance of the student teachers who received traditional tutoring on the occasion of lesson 1 to be approximately the same as the teaching performance of those who did not receive this treatment. On the other hand, the student teachers who received no tutoring were judged to be less proficient in their teaching in lesson 2 than those who were given this treatment.
2. On the question of the behavior of the pupils towards one another, their behavior with the student teachers who received traditional tutorship was significantly more disciplined than with those who had not.

Implications, based on these results

1. The tutor has had a positive influence on the behavior of the student teachers.
2. Traditional tutorship requires the presence of the tutor in the classroom, which could have had the effect of subduing the activity of the pupils (cf. 1:3 above).

4:4 *Results: Student teachers' self-assessment and educational experts' assessments*

The majority of the significant interaction effects also fulfil our requirements for an effect size. Both the effect size of (1) the ATH interaction within the ego-pupil relation and (2) the AUT interaction within the ego-NPO relation fulfil these requirements for the ego dimension. For the pupil dimension the requirements are fulfilled by: (1) the UH interaction within the pupil-pupil relation, (2) the ATH interaction within the pupil-NPO relation, and (3) the AUH interaction within the pupil-NPO relation. Hence the objectivity of the student teachers' self-assessment seems to vary significantly in the following respects: Externally mediated self-confrontation via CCTV/VR has led to significant deviations within the ego-NPO relation, while traditional tutoring has led to demonstrable deviations within the pupil dimension (variable domains 4, 5, and 6).

Implications, based on these results

1. The deviation in the objectivity of the student teachers' assessment of the ego-NPO relation increases for both groups from micro-lesson 1

to micro-lesson 2. The simple effects are not significant, however. The results imply that the student teachers expected an improvement in their teaching method in lesson 2, and that this expectation was felt to be born out irrespective of the actual "objective" state of affairs. On the other hand, the student teachers in the group not having externally mediated self-confrontation via CCTV/VR, have not had the same opportunity of seeing their expectations confirmed in the TV monitor.

2. The deviation in the objectivity of the student teachers' assessment of the pupil-pupil relation shows for both groups an increase in the second lesson. If the comments of the tutor are the source of this increase, there seems to be a disagreement in outlook between the tutor and the panel of experts.
3. The deviation in the objectivity of the perception of the pupil-NPO relation by the student teachers receiving traditional tutoring decreases, while that of those not receiving this treatment increases. For the deviations in evaluation the opposite is the case. One possible explanation could be that the TV monitor has had a standardizing effect on perception, while self-confrontation via the CCTV/VR technique has led to greater deviation in evaluation. The simple effects, however, are not significant within the pupil-NPO relation.

To summarize. The account of the results given above shows that the experimental conditions have not on the whole led to main effects that are significant on the chosen level of significance or that can be regarded as constituting a conclusive basis for interpretation. Taking the third analysis of results into account, this means that there are no deviations in the objectivity of the student teachers' self-assessment as a consequence of either traditional tutoring or externally mediated self-confrontation via CCTV/VR.

On the other hand, examination of the interaction between the experimental conditions and the other sources of variation included in the analysis of variance has produced a number of significant effects that fulfil our criteria for interpretation.

The interaction effects that have been examined more closely imply, for example, that traditional tutorship in the form of dyadic confrontation (Factor H) in combination with perception and evaluation (Factor A) and/or teaching occasion (Factor U) has led to changes in the pupil dimension (variable domains 4—6). This result has emerged in both the student teachers' self-assessment and in the average assessment of the experts.

On the other hand, analysis 1 and 3 show that externally mediated self-confrontation via CCTV/VR (Factor T) in combination with perception and evaluation (Factor A) and/or teaching occasion (Factor U) has led to changes within the ego dimension (variable domain 3).

34.2 Analysis of structure

In order that we might study the structural connection between the student teachers' self-assessment and the experts' assessments (average assessment), the data were treated by means of canonical correlation analysis (cf. e.g. Tatsuoka, 1971). This technique indicates:

1. whether both sets of data are demonstrably related to one another, and
2. the way in which these sets of variables can be combined so that the correlation between the components is at a maximum.

If, as in the present study, it is a question of examining two relatively large sets of variables from the point of view of their interrelationship, then one is primarily interested in a few linear combinations in each group. The variable combinations with the highest correlations are examined first. Moreover, the model means that the structure can as a rule be described almost completely by the first canonical variables, i.e. with a few uncorrelated linear combinations. In other words, the model leads to the relation between the two sets of variables being reduced to its simplest form. For that reason, the method seems to be particularly suitable for use in explorative studies. For a more detailed discussion and description of the method, see part 4, Chs. 28 and 29.

The canonical analysis has been carried out in three stages:

1. the material was examined to find out if there were any significant bivariate relations at all,
2. then the way in which the different variables have contributed to the relation in question was examined, and
3. finally an attempt was made to give these correlations a meaningful content.

Step 1 showed that there are significantly correlated dimensions or common structures in the student teachers' and educational experts' observation data, which are summarized in Table 79.

As can be seen in the table, the analysis has shown nine significant correlated dimensions in the perception structure, while only three correlations have become significant for the evaluation structure. It is obviously easier to achieve a common structure for perception than for evaluation. But for the second lesson, the perception structure shows fewer interrelated dimensions.

Step 2 showed that the weights in the individual dimensions have not on the whole fulfilled the agreement criteria. Thus no common interpretable index can be constructed for either the perception or the evaluation.

Table 79. Number of significant canonical relations (R_c) for perception (a_1) and evaluation (a_2)

Variable domain	Micro-lesson 1		Micro-lesson 2	
	a_1	a_2	a_1	a_2
1. Ego-ego relation	1	1	1	0
2. Ego-pupil relation	1	0	0	1
3. Ego-NPO relation	2	0	0	0
4. Pupil-ego relation	1	0	1	0
5. Pupil-pupil relation	1	0	1	0
6. Pupil-NPO relation	0	0	0	1

Step 3 showed that the weights within each individual dimension have resulted in different signs, and for that reason no separate and interpretable expert or student teacher indexes could be constructed.

Implications

The results of the canonical analyses carried out on the observation data of the educational experts and student teachers show that there are significantly correlated dimensions in the experts' and student teachers' perception and evaluation. As Cooley and Lohnes (1971, p 169) point out geometrically, the canonical correlation can be interpreted as a measure of the extent to which people occupy the same relative positions in the test space of the first set of variables as they do in the test space of the second set.

Since we are concerned with the relationships between the student teachers' and the pedagogical experts' perception and evaluation respectively, the results in Table 79 show that similarities are evident in the perception area. With respect to the evaluation area, however, there are only three significantly correlated dimensions. This result indicates that there is little association between the student teachers' and the pedagogical experts' evaluation. One consequence of the dissimilarities in structure may be that tutors and student teachers run a considerable risk of misunderstanding one another when they try to discuss separate components of a more complex teaching process.

As far as the research method is concerned, difficulties have also arisen in applying this model to the data in question. Some of the problems are discussed in Bierschenk (1972 c). Another possible approach, which might be more suitable for this particular type of problem, is the development of the canonical correlation analysis model into a hypothesis-testing model, where one only decides upon one weight vector (paired components get the same weights) and then calculates the correlations.

35. Recommendations for further research

The results and implications presented above are based on evaluation of only *one part* of the observation data which have been collected in connection with the experiment described in Chapter 5. In this context the recommendations can be divided into two classes:

- (1) general recommendations for further research that can be based on the results and experiences gained during the experiment, and
- (2) recommendations for further analysis, i.e. a study of questions that are stated in the project's original formulation of the problem and for which the author has already collected data.

35.1 General recommendations

On the basis of the results of this experiment which have been presented and discussed in a summarized form in this report, the following can be stated:

1. The teaching staff at the Malmö School of Education, the student teachers training as grade 4—6 teachers who did not take part in the experiment, and the subjects of the experiment have all experienced and evaluated the experiment and thereby externally mediated self-confrontation via CCTV/VR as being "of importance" for teacher training. In addition, the subjects consider self-confrontation via the CCTV/VR technique to be "important" for personality development (one of the goals of teacher training). This result is a positive evaluation of both the medium and the technique of self-confrontation.

It must be considered an extremely important goal for future teacher training to give the student teachers the opportunity of studying and experimenting with their own behavior under systematic and controlled conditions in order to be able to develop different teacher roles and to follow the development of the teaching process.

2. In order to be able to improve the opportunities for research and further investigations into the dimensions which have been studied, the establishment of "Micro-Lesson Laboratories" (MLL) is recommended. MLL

would make possible both individualization and an increased number of training lessons.

As has been made clear in the results presented above, the interaction of the experimental conditions (T, H) with the lesson occasions (U) resulted in significant *and* interpretable effects. On the basis of these results, it can be implied that a longer experimental period and more lessons might very well produce a more definite result. If the student teachers were allowed to give a training lesson regularly, every week for example, it would be possible to make a controlled examination of a greater number of variables in the teaching process than the present experiment has permitted.

3. The development of an MLL system would also facilitate a systematic training of student teachers (and other categories of people) in receiving and processing first-hand information, i.e. non-verbally mediated "self"-information. The experimental data imply that micro-lessons, mediated via the CCTV/VR technique, tend to standardize the perception of student teachers and educational experts.
4. MLL should be equipped with a sufficiently large number of video-recorders and video-tapes to make it possible to store video-recorded micro-lessons for a fairly long period of time. This storage would enable student teachers to re-assess regularly the teaching process in question. The experimental data imply that the student teachers who are given the opportunity of seeing their own lessons via CCTV/VR change their evaluation structure. Being able to get the process of evaluation under systematic control must surely be a very essential goal for teacher training in the future.
5. The results of the structure analyses imply that it is important to investigate in more detail the process of evaluation. A study of the development of the student teachers' perception and evaluation structures should be placed in the centre of future research into CCTV/VR techniques, primarily because the null hypothesis for the main effects in the experimental factors has been accepted.
6. A further study should be made of the effects of the tutoring process on student teachers, since it is not possible to establish within the framework of the experiment whether the observed effects depend simply upon the presence of the tutor in the classroom or upon the influence of the dyadic confrontation.
7. Some experimental results appear to confirm the hypothesis that the expectations of the student teachers direct what is observed in a teaching situation and that these expectations via the TV screen are felt to be corroborated whatever the actual "objective" state of affairs is. Keeping this in mind, a closer examination should be made of the degree to

which the student teachers' predispositions influence the processes of both perception and evaluation. (The author has data available for study of this question.)

8. In order to increase the validity of the measuring instrument, follow-up studies are needed and an extension of the experiment to schools outside the School of Education.
9. The tutor in the experiment has been *one* lecturer in Methodology. In addition he has a particular interest in educational psychology and cannot, therefore, be regarded as being quite representative of the body of lecturers as a whole. Thus no generalizations should be made from the experimental data as far as lecturers in teaching method are concerned. The way in which MLL techniques could be integrated into the School of Education's courses in teaching method requires further studies.
10. Systematic and controlled studies of personality-psychological and social-psychological dimensions in the teaching process require not only new approaches in research method such as the micro-teaching techniques and the CCTV/VR system, but also new approaches in psychometrics, such as the development of statistical-mathematical models that can deal with complex problems. As is implied in the discussion of the separate parts of the analysis of results, the best experimental design and new methods of observation are of little use if the statistical analyses are unsuitable. Thus work is needed here on the development of new evaluation methods.

35.2 Continued tasks of analysis

The experiment presented above is exploratory, which means that a sizable amount of data has been collected in order that the problem might be studied from several different aspects.

The analysis of the data was primarily based on the assessment and evaluation schedule F III. But the test evaluation and coding work have also been completed for all the other data. These data have been stored on magnetic tape, ready for continued analysis. A few examples of such tasks of analysis are given below:

1. Analysis of student teachers' comments

The comments made by the student teachers during the process of self-confrontation were recorded on tape and then worked over in an analysis of content. The coder agreement in coding the physical units, e.g. information units, has been checked.

After this check, recordings were made of both simultaneous comments and comments made during dyadic confrontation, i.e. what the student teachers

and the tutor said during the traditional tutoring. These data have been treated by means of frequency statistics. One of the aims of the evaluation is to examine to what extent the cognition is (1) ego-centered, (2) pupil-centered and (3) topic-centered.

In addition, the analysis of the tutoring comments permits a more systematic examination of the dyadic process of confrontation. The tutor plays a central role in teacher training, and for that reason a more detailed analysis of the pattern of "face-to-face" communication is an important research task.

2. Analysis of the student teachers' reactions to repeated confrontations with one and the same micro-lesson

Assuming that repeated confrontation experiences with a single video-recorded lesson should influence the student teachers' perception and evaluation of the teaching process, their micro-lessons have been played back to them three times. The analysis carried out on the assessment and evaluation schedule F III has resulted in an ANOVA, the examination of which is not yet fully completed, however.

3. Long-term follow-up of student teachers' self-assessment

At the end of their second term at the School of Education (six weeks after the completion of the experiment) and at the end of their period of training at the School (sixth term), the student teachers have been asked to assess once again the micro-lessons video-recorded during the experiment. The test evaluation and coding work have been completed. In addition, these data are stored on magnetic tape. An analysis of these data would aim at studying the extent to which the teacher training has had any effect on the student teachers' perception and evaluation of the micro-lessons video-recorded during the second term.

4. Analysis of measuring instrument

The assessment and evaluation schedule F III has been the main instrument in the experiment but this, like some other schedules included in the group-test battery, is a new construction developed especially for this experiment. To enable the construct validity of the instrument to be studied, a series of factor analyses (cf. Ch. 10) have been carried out for both schedule F III and schedule F II (Identification experiences). As far as schedule F III is concerned, the analytical work has been completed, but it has not yet been reported. An evaluation of schedule F II, e.g. through correlational studies, is also an important task if one wishes to study perception and evaluation tendencies specific for one individual.

5. Analysis of the influence of student teachers' predispositions on their perception and evaluation of teaching processes mediated via video-recording

One interpretation of the results presented above has been that the individual's degree of satisfaction with his own performance before seeing the recording decides to some extent what he will observe on the TV screen, in what way he will evaluate it and what changes it will cause in his attitude.

The analyses of results presented above have been carried out with a view to discovering possible differences between the experimental groups. They imply that one also ought to carry out analyses on the level of the individual, e.g. an analysis of the connection between the experimental results and different personality variables. A group-test battery was administered (see Ch. 8.2.6) for the purpose of showing to what extent the student teachers' perception and evaluation of their own teaching was directed or influenced by the individual's (1) cognitive ability, (2) ability to maintain emotional balance, (3) access to adequate social behavior, (4) ability to use pupil-adapted (concrete) language, (5) ability to stimulate and control the teaching process, (6) ability to maintain opinions despite different types of provocation, (7) ability to achieve an integrative behavior, (8) ability to accept himself and others, (9) ability to make perceptual analysis, and (10) ability to maintain a high level of energy and attention.

The evaluation and coding of separate tests and schedules included in this battery have been completed and these data are stored on magnetic tape.